



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Understanding in vivo modelling of depression in non-human animals: a systematic review protocol

**Citation for published version:**

Bannach-brown, A, Liao, J, Wegener, G & Macleod, M 2017, 'Understanding in vivo modelling of depression in non-human animals: a systematic review protocol', *Evidence-based Preclinical Medicine*, vol. 3, no. 2, pp. e00024. <https://doi.org/10.1002/ebm2.24>

**Digital Object Identifier (DOI):**

[10.1002/ebm2.24](https://doi.org/10.1002/ebm2.24)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Evidence-based Preclinical Medicine

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## SYSTEMATIC REVIEW PROTOCOL

# Understanding *in vivo* modelling of depression in non-human animals: a systematic review protocol

Alexandra Bannach-Brown<sup>1,2</sup>  | Jing Liao<sup>2</sup> | Gregers Wegener<sup>1</sup>  | Malcolm Macleod<sup>2</sup> 

<sup>1</sup>Aarhus Universitet, Aarhus C, Denmark

<sup>2</sup>Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh UK

## Correspondence

A. Bannach-Brown, Clinical Neurosciences, University of Edinburgh, Edinburgh, UK.  
Email: a.bannach-brown@ed.ac.uk

## Funding information

National Centre for the Replacement, Refinement & Reduction of Animals in Research, Grant/Award number: (NC/L000970/1).

The aim of this study is to systematically collect all published preclinical non-human animal literature on depression to provide an unbiased overview of existing knowledge. A systematic search will be carried out in PubMed and Embase. Studies will be included if they use non-human animal experimental model(s) to induce or mimic a depressive-like phenotype. Data that will be extracted include the model or method of induction; species and gender of the animals used; the behavioural, anatomical, electrophysiological, neurochemical or genetic outcome measure(s) used; risk of bias/quality of reporting; and any intervention(s) tested. There were no exclusion criteria based on language or date of publication. Automation techniques will be used, where appropriate, to reduce the human reviewer time. Meta-analyses will be conducted if feasible. This broad systematic review aims to gain a better understanding of the strengths and limitations of current approaches, models and outcome measures used. This study aims to provide insights into factors affecting the efficiency of model induction and the efficacy of intervention. Here, we outline the protocol for a systematic review and possible meta-analysis of the preclinical studies modelling depression-like behaviours and phenotypes in animals.

## KEYWORDS

animal models, depression, risk of bias, validity

## 1 | BACKGROUND

### 1.1 | What is already known about this disease/model/intervention? Why is it important to do this review?

Depression is a mental illness characterized by “*low mood, loss of interest and pleasure or loss of energy.*”<sup>1</sup> It is the leading cause of disability in the world<sup>2</sup> and is currently the brain disorder with the highest financial cost in Europe.<sup>3</sup> The number of people diagnosed with depression worldwide is estimated to be 400 million.<sup>4</sup> Depression places a huge burden on patients and poses a great cost to healthcare systems and governments. The rate of remission with antidepressant medication is, at best, 70% and may only be achieved after several levels of intervention.<sup>5,6</sup> Despite decades of investigation into depression, little is known about the biological mechanisms underpinning the disease.<sup>7,8</sup> With better understanding of the mechanisms

causing depression, the development of novel and more reliable treatments might be possible. There is solid rationale that further investigation into the mechanisms and factors that contribute to the development of depression is needed. This is a highly important area to tackle, both from a clinical and a preclinical perspective.

Preclinical investigations contribute significantly to understanding the mechanisms underlying depression, which can, in turn, inform treatment development and increase translational success of clinical research. One example of this contribution is the systematic review and meta-analysis of antidepressants for the treatment of stroke.<sup>9</sup> Analysis of the evidence in the preclinical animal literature informed key aspects of the trial design in the subsequent FOCUS trial.<sup>10</sup> Preclinical experiments have the ability to model and dissect important mechanisms of depression and therefore provide insights into the neurobiology of the disorder.<sup>11</sup> Preclinical experiments can also investigate the safety and efficacy of proposed treatments prior to exposure in human cohorts.<sup>12</sup> The knowledge from preclinical

investigations can aid prevention research, translating findings into the best and earliest interventions for the human disease, which are top research priorities recently identified by an MQ: Transforming Mental Health report.<sup>13</sup>

Due to the sheer volume of preclinical investigations of depression, it is difficult to achieve an overview of what is already known and to assess the marginal contribution of new research.<sup>23</sup> In this context, a systematic review of the existing preclinical literature could provide an unbiased, collective overview of existing knowledge and allow the additional contribution of new research to be assessed. It could also provide better understanding of the laboratory methods used to induce the condition, the range of outcome measures used to assess depression phenotypes and the variables that might impact the efficacy of different treatments.<sup>14</sup> The findings from this systematic review and meta-analysis may also contribute to the refinement of methods used in animal investigations of depression, reducing the distress caused to animals by substitution with equally informative methods of lower severity, and contribute to the optimisation of the numbers of animals used in depression research by informing well-founded power calculations.

What is meant by an experimental non-human animal model of depression? An experimental model in preclinical non-human animal neuropsychiatric research is defined as including both a dependant variable (an outcome measurement) as well as an independent variable (model induction or manipulation).<sup>15</sup> We differentiate between 3 broad experimental designs within the animal depression literature and anti-depressant drug literature:

1. Studies which compare a control group to a group of animals that receive a lesion (model induction) on an outcome measure. These studies may also have a drug arm.
2. Studies which only compare a "lesion" group of animals to a group of lesioned animals that receive a drug intervention. Once a lesion method has been sufficiently established (known to be valid and reliable), experimenters know a lesion will induce a depressive-like phenotype.
3. Studies which use an outcome measure to assess animals who receive a drug intervention. Once an outcome measure has been established (known to be valid and reliable), experimenters know an outcome measure can reliably measure a depressive-like phenotype.

As a starting point, in order to be thorough, we will look at experiments that investigate the differences between a control group and a lesion or model group. This will provide a basis for characterising a depression model. In these experiments, investigators are directly manipulating a variable intended to produce a depressive-like phenotype and measuring the effects of this manipulation on a given outcome measure. These experiments may or may not include the presence of a drug group/arm.

On this basis, we will characterize all the known models/lesions. From this, drug interventions that have been tested on known models can be characterized. Secondly, the known outcome measures will be extracted from the control vs model investigations. Once the most commonly used outcome measures are known, there is further scope

for characterizing the studies that investigate drug interventions with known outcome measures. We aim to unpick these different experiment design types and evaluate the evidence from all of these (Table 1).

**TABLE 1** Stage of the project at time of protocol submission

| Stage of process                            | Started | Completed |
|---|---------|-----------|
| Preliminary searches                        | Yes     | Yes       |
| Piloting study selection                    | Yes     | Yes       |
| Formal screening with final search criteria | Yes     | Yes       |
| Data extraction from included papers        | No      | No        |
| Quality assessment                          | No      | No        |
| Data analysis                               | No      | No        |
| Manuscript writing                          | No      | No        |

## 2 | OBJECTIVES OF THIS SYSTEMATIC REVIEW

### 2.1 | Specify the disease/health problem of interest

We will investigate how depression is modelled *in vivo*. Depression is defined as specified in the DSM-IV-R under the clinical diagnoses of "Depression," "Major Depressive Disorder (MDD) (Single Episode or Recurring)," "Dysthymia (Persistent Depressive Disorder)," "Depressive Disorder Due to General Medical Condition," "Other Specified Depressive Disorder" and/or "Unspecified Depressive Disorder." This includes depression at any life stage in any gender. In preclinical modelling of depression, some methods induce depressive-like behaviour as a single manifestation rather than modelling the full range of features associated with the clinical diagnoses. Therefore, any model that attempts to mimic one or several major symptoms of depression in an animal model will be considered.

Not all aspects of the human condition can be modelled. Some typically modelled phenotypes in non-human animals include anhedonia and disturbances in sleep and/or food consumption. However, we will not exclude the possibility that novel research has attempted to investigate other aspects not previously modelled in non-human animals, and therefore, there are no restrictions on what phenotypes are modelled, only that they are present in the manifestation of the condition in humans.

### 2.2 | Specify the population/species studied

All preclinical studies on any animal species at any stage of development will be included.

### 2.3 | Specify the intervention/exposure

This study will investigate any mode of inducing depressive behaviour or a model that seeks to mimic the human condition or symptoms of depression using genetic, surgical, pharmacological, developmental or behavioural interventions or a combination of interventions. We will include models induced acutely, chronically, genetically or through a combination of these methods. We will also consider experiments where the efficacy of a treatment or intervention is tested in such models.

## 2.4 | Specify the control population

Studies will be included in this review if they include a suitable control, defined as a cohort of animals that have not been exposed to the method of inducing depressive-like behaviour that was used to create the depressive model. The control cohort may have received an appropriate equivalent, for example, sham surgery instead of lesion or placebo without the active ingredient. For studies that investigate treatment efficacy, a suitable control is defined as a cohort of animals that have had the same exposure to model the disorder as those that are given a treatment but has not been exposed to the treatment tested and may instead receive a placebo in an equivalent route of administration. For studies investigating drug intervention on an outcome measure, a suitable control is defined as a cohort of animals that are not exposed to the drug treatment and may instead receive a placebo in an equivalent route of administration.

## 2.5 | Specify the outcome measures

### 2.5.1 | Primary outcome measure

The primary outcome measure is behavioural outcome measures of animal studies inducing depressive-like phenotype.

### 2.5.2 | Secondary outcome measures

Secondary outcomes include anatomical outcomes, electrophysiological outcomes, neurochemical outcomes and prevalence of reporting of measures to reduce risk of bias.

### 2.5.3 | Tertiary outcome measures

Tertiary outcomes include drug efficacy; inter-rater agreement in the application of the inclusion criteria; and sample size, genomic, proteomic and metabolomic outcomes.

## 2.6 | Research Questions

1. How are animal models of depression induced?
2. What type of outcome measures are assessed in animal models of depression?
3. How precise and accurate are the outcome measures at assessing induced behaviours?
4. To what extent are the outcomes measured in animal models relevant to the endpoints investigated in human trials?
5. How efficacious are different drug interventions in reducing observed manifestations in *in vivo* animal models?

## 3 | METHODS

### 3.1 | Search and study identification

#### 3.1.1 | Identify literature databases to search

Both PubMed and Embase will be searched.

#### 3.1.2 | Define electronic search strategies

See attached for PubMed search terms (Appendix S1, Supporting Information) and Embase search terms (Appendix S2). The animal search filter used for both PubMed and Embase search strings was developed by the

Systematic Review Centre for Laboratory Animal Experimentation, Radboud University Medical Centre (SYRCLE). Reference: A step-by-step guide to systematically identify all relevant animal studies. Marlies Lee-naars, Carlijn R Hooijmans, Niek van Veggel, Gerben ter Riet, Mariska Leeflang, Lotty Hooft, Gert Jan van der Wilt, Alice Tillema, Merel Ritskes-Hoitinga. *Laboratory Animals*. Vol 46, Issue 1, pp. 24 - 31. First published date: January-01-2012. 10.1258/la.2011.011087

#### 3.1.3 | Identify other sources for study identification

Relevant recent reviews will be identified via an additional PubMed search, and the reference list will be searched for any primary research articles that were not identified with the search.

## 3.2 | Study selection procedure

### 3.2.1 | Define screening phases (e.g. pre-screening based on title/abstract, full-text screening, both)

PubMed and Embase search results will be downloaded to EndNote or Reference Manager 12; duplicates will be removed and the full text of articles retrieved where available using the inbuilt feature.

Screening Phase 1: Title and abstracts retrieved from PubMed and Embase will be screened.

Screening Phase 2: Full-text papers will be screened concurrently with data extraction.

### 3.2.2 | Specify number of observers per screening phase: method of screening

#### 3.2.2.1 | Phase 1

A machine learning approach is proposed to assist with the screening phase for inclusion and exclusion criteria. A seed set of papers will be screened by 2 independent human screeners upon which the machine learning algorithm can be trained. Any discrepancies will be resolved by a third human screener. We will pilot the most promising approaches in the context of an ongoing collaboration where we are developing machine learning tools for systematic review. The protocol for this approach is under development and will be uploaded to the CAMARADES website (camarades.info).

Pilot testing of the machine learning algorithm: A random sample of 2000 papers from the overall 70 365 studies identified with the search string was screened for suitability of inclusion by 2 independent reviewers. The decision from the 2000 papers (included or excluded) were used, along with title and abstract, as training sets for several different machine learning algorithms developed by collaborators in the Systematic Living Information Machine (SLIM) consortium. Based on the preliminary sensitivity and specificity analyses and data from a similar project testing the same algorithms using a neuropathic pain dataset,<sup>16,17</sup> an estimated Work Saved over Sampling at 95% recall level ( $WSS@95\% = (TN + FN/N) - 0.5$ ) of above 50% will easily be achieved. Using this approach can reduce the screening workload by at least an estimated 50%, reducing the number of papers needed to be screened by 2 independent human reviewers to less than 35 183 papers.

Quality assessment: A small sub-section of the papers, included and excluded papers, that the machine learning algorithm classifies will be checked by a human screener to ensure the performance of the algorithm.

Validation: We will validate the machine learning techniques for screening by sampling; as opposed to having 2 human screeners manually screen every record. A randomly selected proportion of records that are included and excluded by the algorithm will be double checked by human screeners to ensure the gold standard is maintained. We will continuously monitor the articles screened by the machine learning algorithm by sampling. The machine learning approaches must reach comparable levels to human screening gold standard of at least 95% sensitivity, after which the machine learning algorithm that is maximized for specificity will be chosen.

### 3.2.2.2 | Phase 2

Two independent screeners are responsible for full-text analysis and data extraction, with the aid of machine learning and text mining tools where appropriate, for example, risk of bias classification. A third independent screener will resolve any discrepancies.

## 3.3 | Study selection criteria

### 3.3.1 | Type of study design

#### 3.3.1.1 | Inclusion criteria

Any article providing primary data of an animal model of depression or depressive-like phenotype with an appropriate control group (specified above).

#### 3.3.1.2 | Exclusion criteria

Review article, editorials, case reports, letters or comments, conference or seminar abstracts, studies providing primary data but not appropriate control group.

### 3.3.2 | Type of animals/population

#### 3.3.2.1 | Inclusion criteria

Animals of all ages, sexes and species, where depression-like phenotype intended to mimic the human condition have been induced. Including animal models where depressive-like phenotypes are induced in the presence of a comorbidity (e.g. obesity or cancer).

#### 3.3.2.2 | Exclusion criteria

Human studies and *ex vivo*, *in vitro* or *in silico* studies. Studies will be excluded if authors state an intention to induce or investigate only anxiety or anxious behaviour. Studies will be excluded if there is no experimental intervention on the animals (e.g. purely observational studies).

### 3.3.3 | Type of intervention

#### 3.3.3.1 | Inclusion criteria

All studies that claim to model depression or depressive-like phenotypes in animals. Studies that induce depressive behaviour or model depression and that also test a treatment or intervention (prior or subsequent to model induction), with no exclusion criteria based on dosage, timing or frequency.

#### 3.3.3.2 | Exclusion criteria

Studies that investigate treatments or interventions, but no depressive behaviour or model of depression is induced (e.g. toxicity and side-effect studies).

### 3.3.4 | Outcome measures

#### 3.3.4.1 | Inclusion criteria

Studies measuring behavioural, anatomical and structural, electrophysiological, histological and/or neurochemical outcomes and where genomic, proteomic or metabolomic outcomes are measured in addition to behavioural, anatomical, electrophysiological, histological or neurochemical outcomes.

#### 3.3.4.2 | Exclusion criteria

Where metabolic outcome measures are the primary outcome measure of a study. Where genomic, proteomic, metabolic or metabolomic outcomes are the sole outcome measures in a study, they will be excluded.

### 3.3.5 | Language restrictions

#### 3.3.5.1 | Inclusion criteria

All languages (using automated translations where required).

#### 3.3.5.2 | Exclusion criteria

None.

### 3.3.6 | Publication date restrictions

#### 3.3.6.1 | Inclusion criteria

All publication dates.

#### 3.3.6.2 | Exclusion criteria

None.

### 3.3.7 | Other

#### 3.3.7.1 | Inclusion criteria

Studies must investigate methods or models that induce depressive phenotype/s *in vivo*, or authors must claim that they investigate a model of depression.

#### 3.3.7.2 | Exclusion criteria

Studies claiming to induce only anxiety behaviour or a model of anxiety. In cases where both models of anxiety and depression are investigated, the study will be included, and only the depression-related data will be extracted. In the case of data duplication (2 or more papers reporting the same data), the paper reporting the smallest dataset or fewest outcomes will be excluded. Studies will be excluded if they model aspects of bipolar disorder, manic symptoms, obsessive-compulsive behaviours, panic disorder or psychotic symptoms.

### 3.3.8 | Order of priority exclusion criteria per screening phase

Selection phase 1: screening based on title and abstract

1. Article must be primary research article (excluding reviews, comments or letters).
2. Exclude studies on humans.
3. Exclude *ex vivo*, *in vitro* or *in silico* investigations.
4. Exclude study if no depressive behaviour or model of depression has been induced.

Selection phase 2: full text screening

1. The above criteria from selection phase 1.
2. Exclude if no appropriate outcome is measured.
3. Exclude if no appropriate control group.
4. Where sufficient data cannot be extracted and authors do not respond to requests for required information.
5. Exclude the study with the least information in the case of multiple publications describing the same work.<sup>18</sup>

## 4 | STUDY CHARACTERISTICS TO BE EXTRACTED

### 4.1 | Study meta-data

The first author, corresponding author, year, title, journal name, source of funding and DOI will be extracted.

### 4.2 | Study design characteristics

The number of animals in the experimental and control groups will be extracted. If the number of animals is given as a range, the most conservative estimate will be extracted. The category of experimental design of the study will be extracted (1. Control vs model induction, 2. Model vs drug, 3. Testing drug on outcome measure).

### 4.3 | Animal model characteristics

The species, strain, sex (male or female), age and/or weight of animal will be extracted.

### 4.4 | Method of model induction/intervention characteristics

For studies that induce a model of depression, the method used to induce depressive-like behaviour will be extracted as well as the duration of the model induction. The category of type of model induction will be extracted (i.e., genetic, surgical, pharmacological, developmental or behavioural model induction or a combination of interventions). We will extract information about whether models were induced acutely or chronically or both. If applicable, the following information will be extracted for the method of model induction: the dosage of intervention given, route of delivery, mode of delivery and how long the intervention was given for. The length of time between model induction and outcome measurement will be extracted as well as the length of time between the model induction, any treatment and outcome measurement. For studies with several methods of model induction given, data from all time points will be extracted along with details of each method of model induction.

The information regarding outcome measures described below will be extracted both from experiments describing the induction of a model of depression and from those describing the efficacy of drugs in such models.

### 4.4.1 | Outcome measures

1. Summary outcome data for each group (mean), including whether variance is reported as SD or SEM and the number of animals per group.
2. Details of the outcome measure (e.g. the sub-type or name of the outcome measure and, e.g. in the case of food restriction, the length of time the animal was restricted for).
3. The number of times the outcome measure was assessed.
4. The number of different outcome measures the animal was tested on.
5. The category of the behaviour or biomarker the outcome measure is measuring (e.g. anhedonia, sleep or weight loss, markers of oxidative stress)
6. Any measures taken before the disease model induction will be extracted. The details of the before-and-after comparison will be extracted.
7. Has an appropriate outcome measure been selected for use? Studies that induce a depression model and investigate the effect of a subsequent drug intervention should select a suitable test to measure an outcome (e.g. an outcome measure that does not rely on the same mechanism/behaviour as behaviour that might be affected by side-effects of a given drug).

### 4.4.2 | Treatment characteristics

The following information regarding the treatments tested will be extracted: the dosage of treatment given, route of delivery, mode of delivery, how long the treatment was given for. The length of time between the administration of the treatment and outcome measurement will be extracted as well as the length of time between the model induction and any treatment given if applicable. This information will be extracted regardless of whether an experiment simply assesses an outcome measure for a given drug treatment or if an experiment has induced a model of depression and tests a drug treatment.

### 4.4.3 | Other

The number of excluded animals will be extracted, and the reason for their exclusion, if reported, will be extracted.

## 5 | RISK OF BIAS & STUDY QUALITY

### 5.1 | Criteria to assess the internal validity of included studies

An adjusted CAMARADES checklist will be used to assess risk of bias, including the following criteria:

1. Publication in a peer-reviewed journal.
2. Reporting of random allocation.
3. Reporting of blinding of the conduct of the experiment.
4. Reporting of blinded assessment of outcome.
5. Use of comorbid animals (refers to animals where depression is investigated in the presence of another medical condition, e.g. stroke or diabetes).
6. Reporting of a sample size calculation.



7. Reporting of compliance with animal welfare regulations.
8. Reporting of a potential conflict of interest.
9. Reporting of exclusions of animals.
10. Whether a study protocol is available dated before the experiments began.<sup>19</sup>

We will report the median number of study checklist items scored and the interquartile range.

## 6 | COLLECTION OF OUTCOME DATA

### 6.1 | Methods for data extraction/retrieval

1. Numerical data will be extracted from the full text of publications (mean, SD or SEM, *P* values (exact *P* -value where possible) and group sample size).
2. In studies where data are presented only graphically, the software Universal Desktop Ruler, or a similar tool, will be used to extract the data into numerical values. For certain PDF presentations, it may be possible to use data mining approaches to extract these data.
3. If any data are missing, the corresponding authors will be contacted.
4. In the absence of a response from authors (we will allow 2 months to reply with a follow-up email sent after the 1st month), data will be excluded from analysis.

If the screeners or extractors consider that 2 sources may describe the same data, we will contact the authors seeking clarification. If we receive no response, we will include only the most recent data source.

## 7 | DATA ANALYSIS/SYNTHESIS

### 7.1 | Data gathering and combination

All data will be gathered and entered in the CAMARADES-SyRF database. We will provide a qualitative summary along with several separate meta-analyses, where feasible.

### 7.2 | How the decision as to whether a meta-analysis is appropriate will be made

Based on pilot analyses of a random sample of 2000 from the overall 70 365 studies identified with the search string, approximately 15% of the total records are expected to be relevant to the research question and included in subsequent meta-analyses. This is similar to previous systematic reviews in models of psychiatric disorders conducted at CAMARADES where about 10% to 15% of the studies were included in the analysis. We expect high heterogeneity between studies due to differences in the study designs; therefore, a meta-analysis is proposed to investigate sources of this heterogeneity.

## 8 | IF A META-ANALYSIS SEEMS FEASIBLE/ SENSIBLE

### 8.1 | Effect measure to be used

Mean, SD or SEM and group sample size will be extracted for all outcome measures for both experimental and control groups to calculate pooled effect size. Where a single control group serves multiple intervention groups, the size of the control group used in the meta-analysis will be adjusted by dividing it by the number of intervention groups it serves. If the number of animals is presented in a range, the most conservative estimate will be extracted (e.g. if presented as  $n = 6-12$ , we will consider  $n = 6$ ). *P* values, exact *P* value where possible, will be extracted from primary analyses between model and control and intervention and model in order to conduct *P*-curve analysis.

Categorical or qualitative information relating to the outcome measures, such as the behavioural measure or the symptom the model is trying to elucidate, will be extracted into a text/comment field or into a form drop-down menu.

A decision will be made once the data has been extracted as to which effect size is the most appropriate to use. As most outcome measures are continuous variables, and outcome measures are not likely to be measured on the same scale, Normalized Mean Difference (NMD) effect sizes will be calculated where possible. This effect size calculation will be used where an appropriate "sham" or "control" group is present<sup>20</sup> or where it is possible to impute the outcome in a "normal" animal. If the data are unsuitable for calculating NMD, Standardized Mean Difference (SMD) will be used. NMD and SMD will be calculated using the equations outlined in Vesterinen and colleagues.<sup>20</sup>

### 8.2 | Statistical model of analysis

The data extracted will in all likelihood cover different species, ages and sexes, as well as different study designs and models of induction. Therefore, the true effect size is likely to differ between studies, and a random effects model will be used.<sup>21</sup>

Statistical analyses will be performed using Stata, Statistical Software (College Station, TX: StataCorp LP).

### 8.3 | Statistical methods to assess heterogeneity

Cochran's *Q* will be used for assessing heterogeneity; *Q* is used to calculate the excess variance ( $Q-k$ , where *k* is the degrees of freedom). A *P* value can be calculated for *Q*, giving an indication of whether all studies share a common effect size ( $P < 0.05$ ) or not ( $P > 0.05$ ).  $I^2$  will be used to report heterogeneity as this describes the proportion of observed variance that reflects true differences in effect size between studies.<sup>18</sup>

### 8.4 | Specify which study characteristics will be examined as potential sources of heterogeneity (sub-group analysis)

Meta-regression will be used to investigate the impact of different study characteristics on the outcome, where the effect estimate

(NMD or SMD) is the dependent variable. Categorical variables will be transformed into dummy variables. Where there are sufficient data, a multivariate meta-regression will be used for both model induction and drug models. At least 10 independent comparisons per covariate investigated are required.<sup>21</sup> If there are insufficient data for multivariate meta-regression, univariate analysis will be used, requiring a total of at least 25 independent comparisons.

## 8.5 | Model induction model

Sub-groups analyses:

1. Species of animals (mice vs rats vs etc. vs all)
2. Sex of animals (male vs female vs mixed)
3. Type of animal model
4. Method of model induction (e.g. developmental, genetic, pharmacological, lesion or combination)
5. The outcome measure(s) investigated (behavioural, electrophysiological, neurochemical, anatomical)
6. Number of times the outcome assessment was measured (once vs several)
7. The time from model induction to time of outcome assessment
8. Randomisation (yes/no)
9. Blinding:
  - a. Allocation concealment (yes/no)
  - b. Assessment of outcome (yes/no)
10. Source of funding (public vs industry)

A separate model will be used to investigate the effect of drug intervention on outcomes.

## 8.6 | Drug model

Sub-group analyses:

1. Drug Treatment or Intervention.
2. Method of model induction (e.g. developmental, genetic, pharmacological, lesion or combination), if applicable.
3. The outcome measure(s) investigated (behavioural, electrophysiological, neurochemical, anatomical).
4. Treatment or intervention dose.
5. Treatment or intervention route.
6. Number of times the treatment or intervention is administered.
7. Time the treatment is given in relation to model induction (investigated separately per treatment, pre- or post-model induction).
8. Time the treatment is given in relation to time of outcome assessment.
9. Randomisation (yes/no).
10. Blinding:
  - a. Allocation concealment (yes/no)
  - b. Assessment of outcome (yes/no)
11. Source of funding (public vs industry)

Sensitivity analyses will be performed to assess how missing data from study characteristics and effect size might have affected the results. This will be presented in the form of a summary table.

## 8.7 | Correction for multiplicity of testing

Where there are more than 2 groups being compared in a univariate model, we will use the Holm-Bonferroni correction for multiplicity of testing.

## 8.8 | Method for assessment of risk of publication bias

Risk of publication bias analyses will be assessed using funnel plot assessment, *P*-curve analysis and Egger's regression. Trim and fill analysis will be used to identify potentially missing studies. Analyses will be carried out using SigmaPlot and STATA software package (StataCorp LP; SYSTAT Software Inc).

## ACKNOWLEDGEMENTS

This research is, in part, supported by a scholarship from the Aarhus-Edinburgh Excellence in European Doctoral Education Project. This research is, in part, funded by a grant from the National Centre for the Replacement, Refinement & Reduction of Animals in Research (NC/L000970/1). We thank members from the SLIM (Systematic Living Information Machine) consortium for their collaboration. We thank Zsanett Babor, Sarah McCann and Hanna Vesterinen for their comments on versions of the protocol.

## Conflict of interest

M.M. is co-editor-in-chief of EBPM.

## REFERENCES

1. National Institute for Health and Clinical Excellence (Great Britain). *Depression in Adults: Recognition and Management*. London, UK: National Institute for Health and Clinical Excellence; 2009.
2. Marcus M, Yasamy MT, van Ommeren M, Chisholm D, Saxena S. Depression: a global public health concern. World Health Organisation; 2012. Retrieved November 5, 2015.
3. Gustavsson A, Svensson M, Jacobi F, et al. Cost of disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol*. 2011;21(10):718-779.
4. World Health Organisation (WHO), The World Bank. Out of the shadows: making mental health a global development priority. World Bank Group/IMF Spring Meeting, Washington DC, April 2016.
5. Geddes JR, Carney SM, Davies C, et al. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet (Engl)*. 2003;361(9358):653-661.
6. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *Am J Psychiatry*. 2006;163(11):1905-1917.
7. Nestler EJ, Gould E, Manji H. Preclinical models: status of basic research in depression. *Biol Psychiatry*. 2002;52(6):503-528.
8. Slattery DA, Cryan JF. The ups and downs of modelling mood disorders in rodents. *ILAR J*. 2014;55(2):297-309.
9. McCann SK, Irvine C, Mead GE, et al. Efficacy of antidepressants in animal models of ischemic stroke: a systematic review and meta-analysis. *Stroke*. 2014;45(10):3055-3063.



10. Mead G, Hackett ML, Lundstrom E, Murray V, Hankey GJ, Dennis M. The FOCUS, AFFINITY and EFFECTS trials studying the effect(s) of fluoxetine in patients with a recent stroke: a study protocol for three multicentre randomised controlled trials. *Trials*. 2015;16:369.
11. Krishnan V, Nestler EJ. The molecular neurobiology of depression. *Nature*. 2008;455(7215):894-902.
12. Kiebertz K, Olanow CW. Translational experimental therapeutics: the translation of laboratory-based discovery into disease-related therapy. *Mt Sinai J Med*. 2007;74(1):7-14.
13. MQ: Transforming Mental Health. Depression: asking the right questions; 2016. [http://b3cdn.net/joinmq/10fbab8ed26626f32e\\_a3m6bjx2e.pdf](http://b3cdn.net/joinmq/10fbab8ed26626f32e_a3m6bjx2e.pdf). Accessed January 01, 2016.
14. de Vries RB, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. A search filter for increasing the retrieval of animal studies in Embase. *Lab Anim*. 2011;45(4):268-270.
15. Geyer MA, Markou A. Animal models of psychiatric disorders. In: Bloom FE, Kupfer D, eds. *Psychopharmacology: fourth generation of progress*. New York, NY: Raven; 1995:787-798.
16. Currie GL, Sherratt N, Colvin LA, et al. Search for studies using animal models of neuropathic pain [Data set]. *Zenodo*. 2015. <https://doi.org/10.5281/zenodo.35448>.
17. Howard BE, Phillips J, Miller K, et al. SWIFT-Review: a text-mining workbench for systematic review. *Systematic Reviews*. 2016;5:87. doi:10.1186/s13643-016-0263-z.
18. Babor Z, Macleod M, Thompson L, Sena S, Nunes-Fonseca C.. Improving our understanding of the in vivo modelling of psychotic disorders; 2014. <http://www.dcn.ed.ac.uk/camarades/files/Updated%20Protocol.pdf>
19. Zwetsloot PP, Jansen LSJ, Végh AMD, et al. Cardiac stem cell treatment in myocardial infarction: protocol for a systematic review and meta-analysis of preclinical studies. *Evid Based Preclin Med*. 2015;2(1):10-15.
20. Vesterinen HMES, Sena KJ, Egan TC, Hirst L, Churolov GL, Currie A, Antonic DW, Howells, and Currie MR Macleod. Meta-analysis of data from animal studies: a practical guide." *Journal of neuroscience methods* 221 (2014): 92-102.
21. Borenstein M, Hedges LV, Higgins J, Rothstein HR. *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd; 2009.
22. de Vries RB, Wever KE, Avey MT, Stephens ML, Sena ES, Leenaars M. The usefulness of systematic reviews of animal experiments for the design of preclinical and clinical studies. *ILAR J*. 2014;55(3):427-437.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Bannach-Brown A, Liao J, Wegener G, and Macleod M, Understanding *in vivo* modelling of depression in non-human animals: a systematic review protocol. *Evidence-based Preclinical Medicine*. 2017;e00024. <https://doi.org/10.1002/ebm2.24>